

COMBINED APPLICATION OF LOGISTIC REGRESSION AND THE ANALYTIC HIERARCHY PROCESS TO THE SUPPLIER SELECTION PROBLEM

Stanulevich Ruslan

*student, Belarusian State University of Informatics and Radioelectronics
Belarus, Minsk*

German Yulia Olegovna

*scientific supervisor, Candidate of Technical Sciences, Associate Professor, Belarusian State
University of Informatics and Radioelectronics
Belarus, Minsk*

Supplier selection is one of the key tasks in procurement management for trading enterprises and online stores

Suppliers are characterized by indicators of different nature: price, time, assortment and service indicators, which makes the task multicriteria [1]. With a large number of alternatives, direct application of T. Saaty's analytic hierarchy process to the full sample becomes labor-intensive and insufficiently informative, since some suppliers are obviously ineffective.

The paper proposes a staged procedure: at the first stage, a logistic regression model is constructed and serves as a preliminary filter of alternatives by the probability of a successful commercial operation; at the second stage, the quality and adequacy of the model are checked; at the third stage, a modified Saaty method is used to rank the selected suppliers by integral utility. This combination of statistical classification and expert multicriteria analysis improves the interpretability of the final decision.

Problem statement. The success of a commercial operation is taken as the dependent variable y : $y = 1$ corresponds to the successful completion of the purchase of construction materials, and $y = 0$ corresponds to an unsuccessful operation.

Four selection criteria are used: x_1 – price attractiveness, x_2 – delivery timeliness, x_3 – assortment completeness, x_4 – service quality. The list of criteria is given in Table 1.

To bring the criteria to a common scale [0; 1], normalization is performed by dividing the original values by 10. For example, for the supplier «BazisSnab», the price of 500 rubles is transformed into $x_1 = 0,500$, and delivery timeliness of 70 % is transformed into $x_2 = 0,700$.

The initial sample contains 20 conditional suppliers.

To calculate the integral assessment, utility functions $\varphi_i(x_i)$ defined on the normalized scale are introduced:

$$\varphi_{1(x_1)} = 1 - x_1 \quad (1)$$

$$\varphi_{2(x_2)} = x_2 \quad (2)$$

$$\varphi_{3(x_3)} = x_3 \quad (3)$$

$$\varphi_{4(x_4)} = \frac{|H(x_4) - H(0)|}{|H(1) - H(0)|} \quad (4)$$

where $H(x_i)$ is the normalized Harrington function.

The decrease in $\varphi_1(x_1)$ is explained by the fact that the criterion specifies the price level: an increase in price reduces the attractiveness of a supplier. For x_4 , the Harrington function is used; it reflects the nonlinear effect of service quality and the saturation of utility in the upper part of the scale. For example, the graph of the utility function for criterion x_1 is shown in Figure 1.

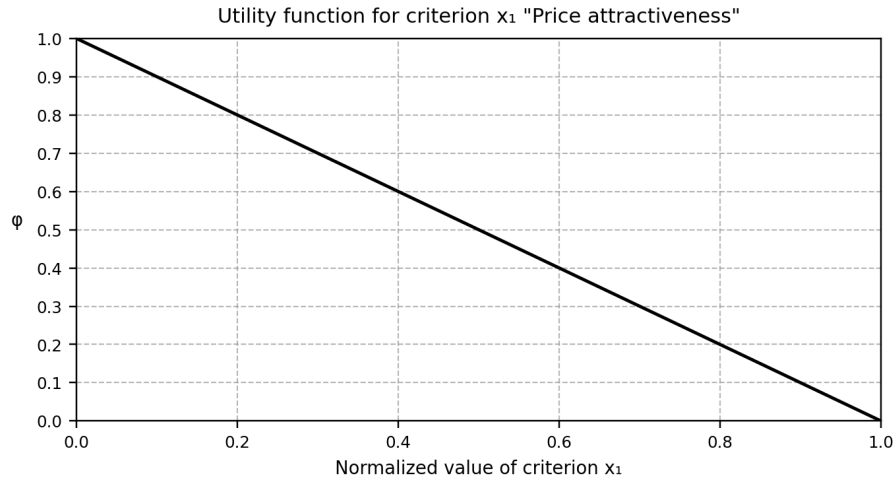


Figure 1. Utility functions for criterion x_1

Construction of the logistic model. Logistic regression in its standard form is written as follows:

$$p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i})}} \quad (5)$$

Using the maximum likelihood method, coefficient estimates were obtained from the sample (Table 1). The features were preliminarily normalized on the scale [0; 1], so the coefficients can be compared by absolute value. The positive sign of b_1 is related to the fact that x_1 specifies the price level: an increase in x_1 reduces the probability of success. For x_2 – x_4 , the signs are negative, which means that an increase in these criteria raises the probability of a successful operation.

Table 1.

Estimates of logistic regression coefficients

Parameter	β	$b = -\beta$
const	–13,7461	13,7461
x_1	–15,7544	15,7544
x_2	21,3570	–21,3570
x_3	17,6434	–17,6434
x_4	10,7390	–10,7390

Model quality assessment. Classification quality was assessed at a probability threshold of 0,5. The share of correct classifications was 0,900, sensitivity was 0,923, specificity was 0,857, AUC = 0,945, and McFadden's pseudo- R^2 was 0,590, indicating a noticeable improvement in the quality of the full model compared with the null model.

Model adequacy was checked using the χ^2 criterion. Suppliers were sorted by the predicted probability p and divided into $k = 10$ groups of two observations each. $\chi^2_{\text{obs}} = 1,705$, $\chi^2_{\text{crit}} = 9,488$ at $\alpha = 0,05$ and $v = 4$, $p\text{-value} = 0,788$. Since $\chi^2_{\text{obs}} < \chi^2_{\text{crit}}$, no statistically significant differences between observed and expected frequencies were found: the model does not contradict the available data and can be used as a preliminary filter.

Application of Saaty's method. At the third stage, the modified T. Saaty method is applied to 13 suppliers assigned by the model to the successful class ($\hat{y} = 1$). The pairwise comparison matrix of the criteria is presented in Table 2.

Table 2.**Pairwise comparison matrix of criteria according to Saaty**

Criterion	x_1	x_2	x_3	x_4
x_1	1	1/2	1/2	2
x_2	2	1	1/2	3
x_3	2	2	1	4
x_4	1/2	1/3	1/4	1

Using the principal eigenvector, the following weights were obtained: $\lambda_1 = 0,182$, $\lambda_2 = 0,286$, $\lambda_3 = 0,435$, $\lambda_4 = 0,097$. The maximum eigenvalue is $\lambda_{\text{max}} = 4,046$, the consistency index $CI = 0,015$, and the consistency ratio $CR = 0,017$. Since $CR < 0,1$, the pairwise comparison matrix is consistent.

The global score G is calculated using the formula:

$$G = \sum_{i=1}^4 \lambda_i \varphi_{i(x_i)} \quad (6)$$

The ranking results for the first six suppliers with the highest value of G are presented in Table 3.

Table 3.**Local utilities and global score G**

Supplier	p	φ_1	φ_2	φ_3	φ_4	G	Rank
BazisSnab	1,000	0,500	0,700	0,900	0,891	0,769	1
StroyMarket	1,000	0,500	0,900	0,600	0,960	0,703	2
ProfKomplekt	1,000	0,400	0,400	0,800	0,960	0,628	3
RegionMaterialy	1,000	0,700	0,300	0,700	0,891	0,604	4
MonolitTrade	1,000	0,500	0,900	0,500	0,387	0,603	5
TekhnoStroy	1,000	0,400	0,800	0,600	0,038	0,566	6

The maximum value of the global score was obtained for the supplier "BazisSnab": $G = 0,769$. Detailed calculation:

$$G = 0,182 \cdot 0,500 + 0,286 \cdot 0,700 + 0,435 \cdot 0,900 + 0,097 \cdot 0,891 = 0,769$$

Conclusions. A two-stage supplier selection procedure combining logistic regression and the modified Saaty method is proposed. The logistic model ensured the selection of «reliable» suppliers, while the Saaty method was applied only to suppliers preliminarily selected by logistic regression, which reduces computational load and improves the interpretability of the result.

In the demonstration sample, the supplier "BazisSnab" received the maximum integral rating ($G = 0,769$).

References

1. Sergeev V.I. Supply Chain Management: textbook for bachelor's and master's students / V.I. Sergeev. – Moscow: Yurait, 2014. – 480 P.
2. Hosmer D.W. Applied Logistic Regression / D.W. Hosmer, S. Lemeshow, R.X. Sturdivant. – 3rd ed. – Hoboken: Wiley, 2013. – 528 P.
3. Aivazyan S.A. Applied Statistics and Fundamentals of Econometrics / S.A. Aivazyan, V.S. Mkhitarian. – Moscow: UNITI, 1998. – 1022 P.
4. Saaty T. Decision Making. The Analytic Hierarchy Process / T. Saaty. – Moscow: Radio i Svyaz, 1993. – 278 P.